控制单水平研究中的多水平误差：基于设计的方法*

王阳 [1]　温忠麟**[2]　付媛姝 [3]

（[1]广东金融学院公共管理学院，广州，510521）

（[2]华南师范大学心理应用研究中心/心理学院，广州，510631）

（[3]肇庆学院教育科学学院，肇庆，526061）

摘　要　由于取样设计的原因，多水平数据结构不仅存在于多水平研究，也广泛存在于单水平研究，需要在单水平分析中控制多水平误差。此时使用多层线性模型发挥不了优势，反而因模型的复杂性带来麻烦。基于设计的方法相对更简单、高效和稳健，更契合含多水平误差的单水平研究情境。在详细介绍基于设计的方法及其优势后，利用数据实例展示基于设计的方法在单水平研究中控制多水平误差的效果，并为应用研究者提供方法选用建议。

关键词　单水平研究；多水平数据；多层线性模型；基于设计的方法

1 引言

在心理、教育和管理等社科研究领域，使用多层线性模型（hierarchical linear model, HLM）分析多水平数据很常见。但少有研究者意识到，单水平研究也会面对多水平数据的问题。最典型的情况是，限于人力和财力成本，许多研究没有采用随机抽样，而是采用整群抽样或多阶段抽样。如在选定的若干学校中，抽取部分或所有学生，这种抽样方式产生的数据往往是多水平结构的（Huang, 2016），即同一学校的学生数据观测值有一定依存性。这时，虽然研究者的目的只是分析学生水平的变量关系，不要分析学校水平的变量（高层级变量），但因回归误差项混入了多水平变异（即多水平误差），若按通常的做法进行最小二乘回归（ordinary least squares regression, OLS; 即我们熟知的单水平回归分析），可能会造成参数估计特别是回归系数标准误估计的偏差（McNeish, 2014a）。文献中将这种研究计划外的多水平结构称为不经意的多水平结构（incidental clustering; McNeish & Wentzel, 2017），以区别于有意设计的多水平结构（deliberate clustering）。

对这种单水平研究遇到多水平数据的情况，比较好的处理方法是什么呢？这个问题还少有研究。尽管可以使用 HLM 来处理，却存在诸多局限。而基于设计的方法（design-based methods, DBM）恰恰是为处理此类问题量身定做的方法，它允许研究者像单水平分析那样处理多水平数据。本文在讨论 HLM 局限的基础上，介绍 DBM 的原理和优势，并通过应用实例展示 DBM 控制单水平研究中多水平误差的

* 本研究得到国家自然科学基金项目（32171091）、广东省哲学社会科学规划项目（青年项目）（GD21YXL04）、广东省普通高校创新团队项目（人文社科）（2019WCXTD005）和广东省教育科学规划项目（2020GXJK342）的资助。

** 通讯作者：温忠麟。E-mail: wenzl@scnu.edu.cn

效果，最后为应用研究者提供方法选用建议。

2 HLM 处理单水平研究中多水平误差的局限

当单水平研究因为取样设计的因素导致数据实际上存在多水平结构时，自然会想到 HLM。但此时使用 HLM 有如下局限。

2.1 优势弱化

如所知，HLM 处理多水平数据实际上包括两种功能：（1）控制多水平误差，保证变量固定效应标准误估计的准确性；（2）分析随机效应并针对特定组的效应做出统计推断。两种功能中，前者并非 HLM 的专利，后者才是 HLM 的独有优势。然而，在以单水平变量关系为关注点的研究中，随机效应是什么样并不要紧，研究任务仅仅是在探讨单水平变量关系的同时控制住多水平误差，即实现功能（1）即可。这样，HLM 的优势并未发挥。而且，回顾以往文献可知，即便在典型多水平研究中，使用 HLM 的目的通常也只是分析固定效应，随机效应多作为厌恶因子（nuisance factor）或次要问题（McNeish, 2014a）。

2.2 缺点放大

单水平研究中，HLM 的优势发挥不出来，缺点却进一步放大了。首先，HLM 对随机效应有多个假设。无论研究者是否关注随机效应，在 HLM 分析时这些假设仍需满足。然而，在单水平研究情境下，由于不是特意的多水平设计，通常层 2 样本容量不足，HLM 分析的有关假设更难满足且难以验证。表 1 总结了违反有关随机效应的假设时，HLM 对层 1 效应（即单水平变量关系）可能产生的估计偏差。虽然这些潜在偏差仅限于具体的模拟情境，不表示任何条件下违反假设都有问题。但至少说明在某些条件下，不满足随机效应假设所带来的风险。

表 1 HLM 涉及随机效应的假设及违反假设可能给层 1 效应分析带来的偏差

HLM 涉及随机效应的假设	违反假设的典型情况	违反假设可能给层 1 效应分析带来的偏差
没有遗漏本应包含的随机效应	只包含随机截距而忽略随机斜率	回归系数标准误被低估，回归系数 95%置信区间覆盖率偏低，第一类错误率偏高
正确定义组内残差的协方差结构	组内残差方差非齐性或残差间有相关	回归系数标准误和回归系数 95%置信区间覆盖率都既可能偏高，也可能偏低
正确定义随机效应的协方差结构	随机效应方差非齐性或随机效应间有相关	回归系数标准误和回归系数 95%置信区间覆盖率都既可能偏高，也可能偏低
组内残差和随机效应不相关	组内残差和随机效应相关系数不为 0	回归系数标准误可能被高估，统计检验力降低
组内残差服从多元正态分布	组内残差不服从正态分布（如偏态分布或 *t* 分布）	回归系数标准误被低估，回归系数 95%置信区间覆盖率偏低，第一类错误率偏高
随机效应服从多元正态分布	随机效应不服从正态分布（如偏态分布或 *t* 分布）	回归系数标准误被低估，回归系数 95%置信区间覆盖率偏低，第一类错误率偏高
预测变量与组内残差无关	预测变量和组内残差相关系数不为 0	回归系数标准误可能被高估，统计检验力降低
预测变量与随机效应无关	预测变量和随机效应相关系数不为 0	回归系数标准误可能被高估，统计检验力降低

注：本表基于下面文献总结得到：Huang（2016）; Jacqmin-Gadda 等（2007）; Litière 等（2007）; McNeish（2019）; McNeish 等（2017）。

第二，为避免因遗漏必要的随机效应而导致估计偏差（McNeish, 2019），有人建议在 HLM 建模时

尽可能多地纳入随机效应（Barr et al., 2013），但这么做也有问题：即使模型理论上可识别，研究者当前数据中的信息也可能难以支撑太复杂的模型（即过参数化, overparameterization），导致模型出现收敛问题，如收敛速度过慢和不收敛等（Bates et al., 2015）。这些问题在各水平样本容量较小的情况下更可能出现。

第三，对于一项单水平研究，使用 HLM 犹如牛刀杀鸡，将问题复杂化。在理论上，需要认真考虑是否加入和加入哪些随机效应；在建模上，需要分别指定组间和组内的变量和模型设定；在结果呈现上，由于数学公式存在嵌套关系（如层 2 截距和斜率方程嵌套于层 1 方程），且需要多重下标来区分变量和误差水平，HLM 经典的符号书写体系远比单水平回归复杂（这直接体现在多水平建模流行软件 HLM 上）；HLM 的结果解释也要比单水平回归复杂。

2.3 小结

当单水平研究因抽样设计的原因导致数据实际包含多水平结构时，用 HLM 进行统计分析的作用仅仅是控制多水平误差，核心优势发挥不出来，却需要研究者承受因 HLM 的复杂性带来的困难。所以，此时 HLM 不是理想的选择。

3 基于设计的方法

既然关注的研究问题仍然是单水平问题，那么，单水平研究者有无可能在自己熟悉的单水平框架下更简单地处理这种实际为多水平结构的数据？DBM 为解决这个问题提供了可能。通常的 OLS 处理多水平数据之所以会出错，是因为回归误差项包含了多水平变异，导致回归系数标准误误估，进而使显著性检验出现偏差。HLM 通过将变量和回归误差分解到不同层级而解决了多水平误差问题；DBM 采用不同策略，即校正而不是分解回归误差项，同样可以准确估计标准误。其本质是控制了多水平误差的单水平回归。这种特殊的单水平回归充分考虑到了复杂抽样设计数据特征（即多水平结构），所以被称为基于设计的方法。

DBM 在近五六年来逐渐受到心理学研究者的关注，不少模拟和实证研究表明它们能有效处理多水平结构的数据（如 Huang, 2016; McNeish & Stapleton, 2016），且相对于 HLM，DBM 有其独特优势，特别是在含多水平误差的单水平研究情境下。通过对各学科多水平分析相关文献的回顾和梳理，可以总结出以下 3 种较值得推荐的 DBM。

3.1 组稳健标准误

第一种常见的 DBM 是组稳健标准误法（cluster-robust standard errors, CRSE），它通过改变回归系数方差（即标准误的平方）的计算方式来校正 OLS 的标准误。对于非嵌套结构的数据，标准的单水平回归可以用方程 $\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon}$ 来表示，其中 $\boldsymbol{Y}$ 是一个 $n\times1$ 的因变量向量，$\boldsymbol{\beta}$ 是一个 $p\times1$ 的回归系数向量，$\boldsymbol{X}$ 是一个 $n\times p$ 的设计矩阵，$\boldsymbol{\varepsilon}$ 是一个 $n\times1$ 的残差向量。回归系数 $\boldsymbol{\beta}$ 的方差可通过下式估计：

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})X(X^{\mathrm{T}}X)^{-1} \tag{1}$$

如果残差项 $\boldsymbol{\varepsilon}$ 是假定独立同分布的（服从均值为 0、方差为 σ^2 的正态分布），则公式（1）可以简化为：

$$\mathrm{var}(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}) = \sigma^2(X^{\mathrm{T}}X)^{-1} \tag{2}$$

计算 $\mathrm{var}(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})$ 对角线元素的平方根，即得到基于 OLS 的系数标准误（McNeish et al., 2017）。

若数据存在多水平结构，残差项包含多水平变异，此时残差方差非齐性，违反同分布假设；同组内残差相关不为 0，违反独立假设。CRSE 分别针对非同分布和非独立问题对标准误进行校正。对于非同分布问题，简式（2）假定方差齐性，会误估回归系数方差，进而误估标准误。因此，CRSE 放弃该简式而采用求取 $\mathrm{var}(\hat{\boldsymbol{\beta}})$ 的完整公式即公式（1）。而对于非独立问题，CRSE 的处理方式是将公式（1）的中间项 $X^{\mathrm{T}}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})X$ 替换为 $\sum_{j=1}^{J} X_j^{\mathrm{T}}E(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^{\mathrm{T}})X_j$，可以理解为各组内个体水平的残差项聚合形成 j 个新的基于组间水平的残差项，用于计算标准误。数据依存性表现在组内水平，而组间水平假定满足独立性假设通常是合理的，所以上述校正解决了独立性问题。有关 CRSE 更多的技术细节可以参考 Cameron 和 Miller（2015）。

使用 CRSE 的主要前提假设是：不同组间的观测数据不相关（McNeish et al., 2017）。这种方法的特点在于，不仅可以用于控制单水平研究中的多水平误差，也可以直接用于分析层 2 水平变量的固定效应，做专门的多水平研究。而且很容易输出 R^2 统计量作为效应量指标，相较之下 HLM 计算效应量要复杂许多。另外，大量的统计软件支持该方法（如 Mplus、R、SPSS、SAS 和 Stata），使其具备了广泛应用的条件。

CRSE 的一个主要不足在于，它仅仅是对回归系数标准误的校正，并不校正回归系数本身。回归系数估计的准确性依赖于单水平回归原本采用的参数估计方法（如 OLS 或极大似然估计）。这在纵向数据或组内相关系数（intraclass correlation coefficient, ICC）极高的数据中可能存在问题（McNeish et al., 2017）。好在至少横截面研究中，即便是 OLS，回归系数估计也是准确的（Huang, 2018; McNeish, 2014a），任何多水平分析方法的主要作用仅仅是准确估计回归系数标准误。这保证了 CRSE 整体上的可靠性。

3.2 广义估计方程

第二种常见的 DBM 是广义估计方程（generalized estimating equations, GEE; Liang & Zeger, 1986）。它的基本步骤是：先由研究者设定一个反映组内观测值相关关系的工作相关矩阵（working correlation matrix）的结构，然后，在假定误差独立的前提下（忽略多水平结构，如同单水平回归那样）拟合回归模型，求得回归系数和残差；再使用上一步得到的残差信息估计出工作相关矩阵的初始值，以符合研究者设定的结构，并用工作相关矩阵估计结果变量（在每组内）的协方差矩阵，更新回归系数和标准误，以反映观测值之间的依存性；通过不断地在更新的工作相关矩阵、结果变量的协方差矩阵以及回归系数

估计值之间迭代，直至回归系数在两次迭代之间不再有明显的变化，模型收敛，回归系数的估计至此结束；最后用前文介绍的 CRSE 来校正标准误。有关 GEE 更多的技术细节可以参考 McNeish（2019）。

使用 GEE 主要的前提假设是：（1）不同组间的观测数据不相关；（2）工作相关矩阵的设定近似于真实结构（McNeish et al., 2017）。该方法不仅像 CRSE 一样，可以直接分析层 2 水平变量，还具有一个其它 DBM 没有的独特优势：适合分析追踪数据和类别结果变量（McNeish, 2014b; McNeish et al., 2017）。这使其成为功能最为全面的 DBM。GEE 的相对局限在于，没有可用的效应量以及拟合评价指标，不利于研究者评价效应大小和模型拟合。

3. 3 固定效应模型

固定效应模型（fixed effects model, FEM）通过将多水平数据中的组别标识变量设定为哑变量纳入回归模型（比如学生嵌套于学校时，可将学校编号设为哑变量），控制了所有层 2 水平变异，从而避免分析层 1 变量间关系时受到误差非独立性的影响（McNeish & Kelley, 2019）。以学生嵌套于学校为例，假定 X 和 Y 分别是学生层面的自变量和因变量，学校共有 3 个，则固定效应模型可以用下式表示：

$$Y = \gamma_0 + \beta X + \gamma_1 S_1 + \gamma_2 S_2 + r \qquad (3)$$

其中，β 是 Y 对 X 的回归系数，它表示对于所有学校而言层 1 自变量对因变量的平均效应，S_1 和 S_2 分别为学校 1 和 2 所代表的哑变量，学校 3 为参照组，γ_0 表示参照组的因变量均分，γ_1 和 γ_2 分别表示学校 1、2 和学校 3 在因变量得分上的均值差异，r 是残差项，仅代表层 1 水平误差。

使用 FEM 主要的前提假设是：残差服从正态分布（McNeish & Kelley, 2019）。该方法最大的优势是，所有层 2 水平变异被完全控制。这带来两个好处：（1）完全消除了 HLM 等其它多水平分析方法可能出现的因遗漏必要的层 2 协变量而导致的估计偏差（McNeish & Kelley, 2019）；（2）对于多水平结构数据，通常要给所有层 1 自变量做组均值中心化（group mean centering, GMC）处理，避免回归系数因混入层 2 效应而不准确（方杰等, 2010）。OLS、HLM、CRSE 和 GEE 都需要这个步骤。但 FEM 由于控制了一切层 2 变异，不用做中心化处理就已经能获得可靠的层 1 回归系数，精简了操作。另外，FEM 应用门槛极低，对统计软件近乎没有要求，任何能做回归分析的软件就能做 FEM。

FEM 的主要缺点在于，它不能像其它 DBM 那样分析层 2 变量，只能分析层 1 效应；其次，FEM 视层 2 抽样单位为固定而非随机的，因此无法将分析结果推论到未抽样的层 2 单位中。不过，在含多水平误差的单水平研究情境，由于层 2 组数过少，即便是 HLM，也很难将结果推论到其它层 2 单位，所以这个缺陷不是大问题。

3. 4 基于设计的方法的共同优势

3. 4. 1 契合含多水平误差的单水平研究情境

当单水平研究含多水平误差时，HLM 分析随机效应的核心优势变得多余甚至累赘。与之不同，DBM 天生不具备分析随机效应的能力，其主要作用就是在单水平研究中消除多水平误差这一“噪音”，保证

层 1 回归系数标准误估计的准确性，既满足了单水平研究的主要需求，又没有功能上的浪费，方法和研究问题更契合。

3.4.2 更少的假设和更大的参数估计稳健性

DBM 不能设定随机效应，使其避免了对随机效应相关假设的依赖，从而减少了因这类假设被违反而面临的风险，具备更好的稳健性。这在含多水平误差的单水平研究中是一个重要优势，因为此时这些假设更难检验且更难满足。比如，McNeish（2019）的模拟和实证研究表明，当遗漏随机斜率和随机斜率非正态时，GEE 都能保持参数估计的准确，而 HLM 则出现层 1 标准误低估问题。由于 CRSE 是 GEE 的一个步骤，在组内相关系数较小的横截面研究中基本等价于 GEE，可以预期 CRSE 在类似的条件下也有较好的稳健性。此外，McNeish 和 Stapleton（2016）的一项模拟研究发现，在层 2 样本容量极少时，HLM 对层 1 回归系数的 95%置信区间覆盖率偏低，系数标准误低估，而 FEM 表现良好。这也说明了 FEM 的相对稳健性。

3.4.3 更少的收敛问题

复杂的随机效应设定给模型的收敛带来更大的麻烦（McNeish & Stapleton, 2016），DBM 不关注随机效应使得由此带来的收敛问题大大减少。不少模拟研究表明，CRSE、GEE 以及 FEM 的收敛率都要高过 HLM （Bolin et al., 2019; Huang, 2018; McNeish, 2019）。实证研究也有相关证据。如在朱瑜和谢斌斌（2018）的一项含多水平误差的单水平中介研究中，作者首先采用 HLM 建模，未能收敛，转而使用 CRSE 就没有出现问题。

3.4.4 模型简洁

DBM 还有一个突出优势在于其模型简洁。DBM 既不需要研究者考虑每个变量究竟属于哪个层级，也不需要像 HLM 一样将模型残差分解到不同水平上，而是把所有层级变量和残差都放在同一个单水平回归模型中。这样，模型基本形式得以简化。比如，对于一个含有层 2 水平误差的单水平中介模型（即 1-1-1 型中介模型; 方杰等, 2010），所建立的 HLM（假定斜率无随机效应）为：

层 1： $$M_{ij} = \beta_{M0j} + \beta_{M1j}(X_{ij} - X_{.j}) + r_{Mij} \quad (4)$$

层 2： $$\beta_{M0j} = \gamma_{M00} + \gamma_{M01}X_{.j} + u_{M0j} \quad (5)$$

层 2： $$\beta_{M1j} = \gamma_{M10} \quad (6)$$

层 1： $$Y_{ij} = \beta_{Y0j} + \beta_{Y1j}(X_{ij} - X_{.j}) + \beta_{Y2j}(M_{ij} - M_{.j}) + r_{Yij} \quad (7)$$

层 2： $$\beta_{Y0j} = \gamma_{Y00} + \gamma_{Y01}X_{.j} + \gamma_{Y02}M_{.j} + u_{Y0j} \quad (8)$$

层 2： $$\beta_{Y1j} = \gamma_{Y10} \quad (9)$$

层 2： $$\beta_{Y2j} = \gamma_{Y20} \quad (10)$$

其中 $X_{.j}$ 和 $M_{.j}$ 是自变量和中介变量的组均值，代表两个变量的组间成分，X_{ij}-$X_{.j}$ 和 M_{ij}-$M_{.j}$ 是自变量和中介变量的组均值离均差，代表两个变量的组内成分；γ_{M00} 和 γ_{Y00} 是中介变量和因变量的截距项，γ_{M10} 和 γ_{M01} 是中介效应前段路径系数 a 的组内效应和组间效应；γ_{Y20} 和 γ_{Y02} 是中介效应后段路径系数 b 的组内效应和组间效应；γ_{Y10} 和 γ_{Y01} 是控制了 M_{ij} 的中介作用后自变量 X_{ij} 对 Y_{ij} 的直接效应 c'的组内成分和组间成分；u_{M0j} 和 r_{Mij} 是 M_{ij} 的层 2 及层 1 残差项；u_{Y0j} 和 r_{Yij} 是 Y_{ij} 的层 2 及层 1 残差项。

上述方程组共包含 7 个方程，17 种系数，方程存在嵌套关系，且所有符号至少双重、最多三重下标（这是因为至少需要用符号 i 和 j 来表示层 1 个体和层 2 组织），如果还想加上随机斜率，模型将进一步复杂化。但若采用 CRSE 和 GEE，模型可以简化为：

$$M = \gamma_M + a_w(X - Xm) + a_b Xm + r_M \quad (11)$$

$$Y = \gamma_Y + c'_w(X - Xm) + c'_b Xm + b_w(M - Mm) + b_b Mm + r_Y \quad (12)$$

上述公式与单水平回归的基本形式并无不同，HLM 方程中的 $X_{.j}$、$M_{.j}$、X_{ij}-$X_{.j}$、M_{ij}-$M_{.j}$、γ_{M00}、γ_{Y00}、γ_{M10}、γ_{M01}、γ_{Y20}、γ_{Y02}、γ_{Y10}、γ_{Y01}、u_{M0j}+r_{Mij} 和 u_{Y0j}+r_{Yij} 分别被 Xm、Mm、（X-Xm）、（M-Mm）、γ_M、γ_Y、a_w、a_b、b_w、b_b、c'_w、c'_b、r_M 和 r_Y 替代，由于变量和误差不用分层标记和放置，方程减少到 2 个，系数减少到 10 种，下标最多一级。

如果使用 FEM（假设层 2 抽样单位为 3 个），模型还可以进一步简化：

$$M = \gamma_M + aX + \gamma_{M1}S_1 + \gamma_{M2}S_2 + r_M \quad (13)$$

$$Y = \gamma_Y + c'X + bM + \gamma_{Y1}S_1 + \gamma_{Y2}S_2 + r_Y \quad (14)$$

上述模型不仅不必分层，还因为组别标识哑变量解释了所有层 2 变异，不需要对层 1 变量进行组均值中心化以及将组均值纳入模型，只要使用原始的 X 和 M 分数，就可以准确估计单水平中介效应，比 CRSE 和 GEE 还简单。几种方法对应的中介模型如图 1，可见从左到右模型越来越简洁。

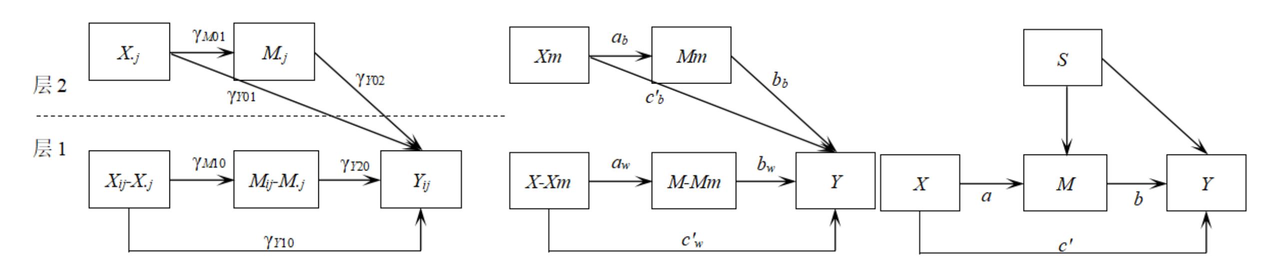


图 1 含多水平误差的单水平中介模型（左：HLM；中：CRSE 和 GEE；右：FEM）

总之，DBM 比 HLM 更简洁直观，模型系数的意义一目了然。研究者只要熟悉单水平回归，就可以和 DBM 无缝衔接，且对分析结果的解释也和单水平回归一样。

3.4.5 操作简单

DBM 建模的简洁直接反映在统计操作上。以 CRSE 为例，该方法在 Mplus 软件中很容易实现，对同一个模型进行分析时，比 HLM 的操作命令简单。首先，HLM 的 analysis 命令需要依研究水平的不同、

关注随机效应的变化而变化，CRSE 只要固定使用“estimator=mlr; type= complex;”这两条命令；第二，在 variable 和 model 部分，多层线性模型都要按变量层级分别设定命令（此部分尤其容易出错），而 CRSE 只需像单水平回归一样设定变量和模型，不必分层；第三，CRSE 能直接用画图功能输出模型图，HLM 则不能。除了 CRSE，其它 DBM 也很容易操作，GEE 和 FEM 都能在 SPSS 中实现窗口化操作（见附录），即通过点击鼠标完成操作，而非编写程序代码。

3.4.6 运行高效

DBM 由于不分析随机效应，软件运行速度均明显快于 HLM。HLM 运行时间大大依赖于随机效应、收敛标准、样本容量和所用软件。比如，一个包含 3 个随机效应的 HLM 若采用自适应高斯积分求解需 18 个小时收敛，采用线性化方法求解需 8 分钟收敛，而同样的分析使用 GEE 只需 3 秒（McNeish, 2019）。

3.5 小结

在心理学研究领域，DBM 作为一种较新的多水平数据分析方法，使研究者不必跳出单水平研究的逻辑框架这一“舒适区”，即可有效控制多水平误差，准确分析变量关系，且具备假设更少、模型简洁、操作简单、运行高效及稳健性好的特征，在单水平研究情境下较好替代了 HLM。

当然，DBM 并非没有局限。所有 DBM 共同的问题在于，随机效应只能控制不能分析。比如，如果研究者想了解不同高校的教师平均情绪智力与职业倦怠的关系是否有差异，可以通过建立带有随机斜率的多层线性模型来分析这个问题。进一步地，研究者还可以针对特定的学校做统计分析，即每个高校的教师平均情绪智力对职业倦怠的效应具体是多大。对于上述这些问题，HLM 是唯一可用的分析手段，DBM 无能为力，它只能就教师情绪智力与职业倦怠的关系在所有学校的平均水平上做一个整体推断，尽管对单水平研究而言这已经足够。

4 应用实例

下面用一组模拟生成的数据来展示本文介绍的 DBM 控制单水平研究（以 1-1-1 型中介为例）中的多水平误差的效果，并与 OLS、HLM 做比较。根据已有研究（方杰等, 2010），这类中介效应可分解为组间中介（高层级中介效应，见图 1 左图层 2 部分模型）和组内中介部分（低层级中介效应，见图 1 左图层 1 部分模型），后者反映的才是单水平意义上的中介效应，因此我们只关注后者（简便起见组内中介前、后段路径系数直接用 *a* 和 *b* 表示）。系数 *a* 和 *b* 的真值均为 0.30，它们对应的标准误真值分别为 0.029 和 0.041；中介效应 *ab* 真值为 0.09，其标准误真值为 0.016。为较好地模拟单水平研究存在多水平误差的情境（层 2 样本容量很小，层 1 样本容量较大），我们生成的数据层 2 样本容量仅为 3 组，每组 200 人，层 1 总样本容量为 600 人，接近通常的整群抽样效果。因变量 *Y* 和中介变量 *M* 的 ICC 分别是 0.96 和 0.30，数据存在明显多水平结构。

分别用 OLS、预测变量（包括自变量和中介变量）组均值中心化后的 OLS（OLS-GMC）、HLM、CRSE、GEE 和 FEM 分析数据（操作模板见附录）。所有方法，除了 OLS 和 FEM，其它方法均预先对

预测变量做组均值中心化处理。结果表明（见表 2）：除了中介前段路径外，OLS 对中介模型中其它系数和标准误的估计都明显偏高。在对预测变量做组均值中心化处理之后，OLS 估计回归系数的偏差完全消除，但系数 *b* 和 *ab* 的标准误仍然明显偏高。以上结果证明，OLS 如果不对标准误做任何校正，是无法控制多水平误差的。

对于所有 DBM 和 HLM，回归系数估计都是精确的，3 种 DBM 对模型中所有标准误的估计偏差都很小（除了 CRSE 的 *ab* 标准误相对估计偏差稍高于 10%），而 HLM 对所有标准误的估计都偏低。总的来看，参数估计 FEM 表现最佳，GEE 其次，CRSE 又次之，HLM 表现相对最差。

就操作过程来说，FEM 最简单，不仅能通过 SPSS 窗口化操作，还不用像其它方法一样中心化和控制组间变量，只需使用 *X* 和 *M* 的原始分和组别标识哑变量。此外，借助 PROCESS 这一流行的 SPSS 中介调节分析插件，FEM 能同时实现依次检验法和系数乘积法并输出 bootstrap 区间（HLM 给不出 bootstrap 区间），还可以完成很多复杂中介和调节混合模型的窗口化操作。GEE 也可以通过 SPSS 窗口化操作，只是不能直接检验系数乘积。CRSE 和 HLM 都需要编写 Mplus 语句完成，不过 CRSE 不涉及分层操作（即与 between 和 within 有关的命令），比 HLM 操作简单，也更不易出错。

表 2 各方法对中介模型中固定效应及其标准误的估计值

方法	中介前段路径（*a*）		中介后段路径（*b*）		中介效应（*ab*）	
	β	*SE*	β	*SE*	β	*SE*
OLS	0.308^{***}	0.033	2.794^{***}	0.147	0.860^{***}	0.104
OLS-GMC	0.303^{***}	0.027	0.287	0.229	0.087	0.070
HLM	0.303^{***}	0.022	0.287^{***}	0.030	0.087^{***}	0.010
CRSE	0.303^{***}	0.027	0.287^{***}	0.037	0.087^{***}	0.013
FEM	0.303^{***}	0.027	0.287^{***}	0.044	0.087^{***}	0.016
GEE	0.303^{***}	0.030	0.287^{***}	0.044		

注：用 SPSS 中的 GEE 无法直接检验 *ab* 系数乘积。***，$p<0.001$。

5 方法选择建议

以上介绍了可以在单水平研究中控制多水平误差的方法，那么在应用研究中应当如何选择方法呢？基于研究者的实际研究目的和各方法的特点，这里提出以下建议：

（1）考虑到 DBM 更加匹配单水平研究情境，当研究者 HLM 知识基础不足且对随机效应分析没有需求时，推荐优先选用 DBM。而在各种 DBM 当中，考虑到 FEM 参数估计的精确性最高，操作最简单，而且很容易和 PROCESS 结合起来实现各种复杂模型的窗口化分析，建议优先选择 FEM。

（2）如果除了层 1 效应，研究者也想分析一下某些层 2 固定效应，如层 1 自变量的情境效应，那

么FEM无能为力，此时建议优先选择GEE或CRSE，但使用这两种方法时层2样本容量不要太小（Huang, 2018; McNeish & Stapleton, 2016）。

（3）当研究者具备足够的 HLM 基础，且需要探讨随机效应时，应当特意收集多水平数据，特别是保证层 2 样本容量充足。若模型能够收敛，建议研究者使用包含所有可纳入随机效应的 HLM，并以稳健极大似然估计（robust maximum likelihood estimator, MLR）作为参数估计方法，这样能尽可能减少由随机效应假设违反导致的问题，并实现对随机效应的分析。

（4）无论 HLM 还是 DBM，分析数据都要用到组别标识变量。即便是做单水平研究，也建议在收取数据时保留各层抽样单位的组别标识信息，以免数据实际层级超过预设水平，导致无法控制多水平误差，从而降低研究结果的可靠性。

致谢：感谢宁夏大学教育学院王惠惠副教授和江西师范大学心理学院戴步云博士为本文写作提供的帮助。

参考文献

方杰, 张敏强, 邱皓政. (2010). 基于阶层线性理论的多层级中介效应. *心理科学进展, 18*(8), 1329–1338.

朱瑜, 谢斌斌. (2018). 差序氛围感知与沉默行为的关系：情感承诺的中介作用与个体传统性的调节作用. *心理学报, 50*(5), 539–548.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). *Parsimonious mixed models.* Retrieved May 26, 2018, from https://arxiv.org/pdf/1506.04967.pdf

Bolin, J. H., Finch, W. H., & Stenger, R. (2019). Estimation of random coefficient multilevel models in the context of small numbers of level 2 clusters. *Educational and Psychological Measurement, 79*(2), 217–248.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372.

Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *Journal of Experimental Education, 84*(1), 175–196.

Huang, F. L. (2018). Multilevel modeling and ordinary least squares regression: How comparable are they? *Journal of Experimental Education, 86*(2), 265–281.

Jacqmin-Gadda, H., Sibillot, S., Proust, C., Molina, J. M., & Thiébaut, R. (2007). Robustness of the linear mixed model to misspecified error distribution. *Computational Statistics and Data Analysis, 51*(10), 5142–5154.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13–22.

Litière, S., Alonso, A., & Molenberghs, G. (2007). Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics, 63*(4), 1038–1044.

McNeish, D. M. (2014a). Analyzing clustered data with OLS regression: The effect of a hierarchical data structure. *Multiple Linear Regression Viewpoints, 40*(1), 11–16.

McNeish, D. M. (2014b). Modeling sparsely clustered data: Design-based, model-based, and single-level methods. *Psychological Methods, 19*(4),

552–563.

McNeish, D. M. (2019). Effect partitioning in cross-sectionally clustered data without multilevel models. *Multivariate Behavioral Research, 54*(6), 906–925.

McNeish, D. M., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods, 24*(1), 20–35.

McNeish, D. M., & Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research, 51*(4), 495–518.

McNeish, D. M., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*(1), 114–140.

McNeish, D. M., & Wentzel, K. R. (2017). Accommodating small sample sizes in three-level models when the third level is incidental. *Multivariate Behavioral Research, 52*(2), 200–215.

附录 控制单水平中介分析中的多水平误差的软件操作

HLM 的 Mplus 语句

```
DATA: FILE = 示例数据.csv;
VARIABLE: NAMES = id y group x m xc mc xm mm;
!group 组别标识变量；xc、mc 自变量和中介变量的组均值离均差；xm、mm 自变量和中介变量组均值
Usevariables= y group xc mc xm mm;
CLUSTER=group;!指定组别标识变量
WITHIN = xc mc; !指定层 1 变量
BETWEEN = xm mm; !指定层 2 变量
ANALYSIS: estimator=MLR; TYPE = TWOLEVEL;
MODEL: %WITHIN%
y on xc mc; mc on xc; !设定层 1 模型
%BETWEEN%
y on xm mm; mm on xm; !设定层 2 模型
MODEL indirect: y ind xc; !分析中介模型组内成分
OUTPUT: STANDARDIZED(STDYX);
```

CRSE 的 Mplus 语句

```
DATA: FILE = 示例数据.csv;
VARIABLE: NAMES = id y group x m xc mc xm mm;
Usevariables= y group xc mc xm mm;
```

CLUSTER=group;

ANALYSIS: estimator=MLR; TYPE=COMPLEX; !使用 CRSE 的关键指令

MODEL: y on xm mm xc mc; mc on xc; mm on xm; !不分层设定模型

model indirect: y ind xc;

output: STANDARDIZED(STDYX);

FEM 的 SPSS 窗口操作步骤

1、在 SPSS 菜单栏依次选择：转换——创建虚变量，将组别标识变量 group 选入右侧输入框。然后输入哑变量根名称(使用原始变量名称 group 即可)并点击“确定”，然后在生成的组别标识哑变量中任意删去一列。

2、在 SPSS 菜单栏依次选择：分析——回归——PROCESS（默认已安装 PROCESS 插件），将自变量 x、因变量 y 和中介变量 m 分别选入对应的输入框，并将生成的组别标识哑变量选入协变量框（Covariates）。Model number 选择 4（代表简单中介模型），然后“确定”即可。

GEE 的 SPSS 窗口操作步骤

1、在 SPSS 菜单栏依次选择：分析——广义线性模型——广义估算方程。

2、“重复”选项卡：将分组变量 group 选入“主体变量”框，并在“协方差矩阵”单选框中选择“基于模型的估计量”。

3、“响应”选项卡：指定后果变量。分析中介前段路径时将中介变量 m 选入；分析中介后段路径时将因变量 y 选入。

4、“预测变量”选项卡：指定前因变量。分析中介前段路径时将自变量（包括离均差 xc 和组均值 xm，共 2 个变量）选入“协变量”框，分析中介后段路径时将自变量和中介变量（包括变量的离均差 xc、mc 和组均值 xm、mm，共 4 个变量）都选入“协变量”框。

5、“模型”选项卡：将所有预测变量选入“模型”框，然后点击“确定”。

注意：因为是依次检验法，中介前后两段路径要分两次进行操作。

Controlling for Clustering in Single Level Study: Design-Based Methods

Wang Yang[1], *Wen Zhonglin*[2], *Fu Yuanshu*[3]

([1] School of Public Administration, Guangdong University of Finance, Guangzhou 510521, China)

([2] Center for Studies of Psychological Application / School of Psychology, South China Normal University, Guangzhou 510631, China)

([3] School of Education, Zhaoqing University, Zhaoqing 526061, China)

Abstract In social science research fields, single-level research often adopts cluster sampling or multi-stage sampling to obtain samples, resulting in the fact that the data structure is multi-level. Thus, researchers have to control for errors from the higher level in their single-level studies.

Hierarchical linear model (HLM) suffers from limitations in dealing with such issue. First, HLM's unique advantage to

focus on random effects and cluster-specific inferences cannot be reflected in single-level research. Second, the disadvantages of HLM are amplified in single-level research. (1) HLM's assumptions about random effects are harder to satisfy and test. Violation of these assumptions may result in parameter estimation bias. (2) HLM is more likely to produce convergence problems. (3) For single-level studies, HLM is complex in theory, modeling, software operation and interpretation of results. Thus, HLM is difficult to generalize in a single level study with multi-level error.

Design-based methods (DBM), including cluster-robust standard errors (CRSE), generalized estimation equation (GEE), and fixed effects model (FEM), represent a category of logical and valid procedures to analyze multi-level data. By correcting for the standard errors of fixed effects, DBM circumvents the issues of partitioning residuals and variables into different levels while accurately estimate parameters. Thus, DBM can address multi-level data within the single-level framework, which is very friendly to single-level researchers.

Contrast to HLM, DBM is more parsimonious in modeling, simpler in operating, more efficient in running and more robust in estimating for single-level research. Therefore, at least under the condition of single-level research with multi-level error, DBM is an ideal alternative to HLM.

After a detailed introduction of DBM and its advantages, a simulation data set were used to demonstrate the effectiveness of DBM in controlling for multi-level error in single-level mediation studies (i.e., 1-1-1 mediation model). The results showed that although both HLM and DBM were accurate in estimating the within-cluster component of the mediating effect, the former underestimated the standard errors of mediating effect and each mediating path coefficient. In addition, all of the DBMs are simpler than HLM in terms of operations, especially the FEM. FEM is not only possible to operate through SPSS, but also unnecessary to center the variables in level 1 and control between-cluster variables. What's more, through the popular SPSS mediating analysis macro PROCESS, FEM can realize both casual steps approach and coefficients product approach with bootstrap confidence interval for various complex mediation models.

Finally, following suggestions were given for practitioners to select appropriate methods to accommodate clustering in single-level research. (1) DBM is suggested to control the multi-level error in single-level study, especially FEM. (2) If researchers are interested in between-cluster fixed effects, CRSE and GEE is recommended. (3) When researchers have sufficient background knowledge of HLM, and need to focus on random effects, they should collect multi-level data deliberately, especially to ensure that the sample size of level 2 is sufficient. (4) It is recommended to retain the cluster identification information when collecting data, so as to prevent the actual level of data from exceeding the expectant level, leading to the failure to control the multi-level error.

Key words single-level research; clustered data; hierarchical linear model; design-based methods